

Data Management Plan

The HAWC collaboration has a public Data Management Plan. Our plan complies with the NSF requirements for public data release. Beyond this basic requirement, our goal is to also provide easy to use data tools that are useful to other scientists in the field. To this end we have a plan for release of a number of high-level data products described below.

A data management plan for the operation of the HAWC detector has been approved by the DOE and NSF as part of a larger Project Operations Plan (POP). This document summarizes the data management portion of the plan.

Products of the Research: The raw data produced by the HAWC experiment consists of timing and time-over-threshold information for pulses generated by any of 1200 photomultiplier tubes (PMTs) that are hit by Cherenkov light during an atmospheric shower and the rates of these hits recorded by scalars. Since the summer 2018, the pulses from an additional 350 PMTs in the high energy outrigger array are also recorded. Intermediate-level data products include reconstructed air shower directions, core locations, energy, and gamma/hadron likelihood. It also includes simulated PMT hits from Monte Carlo simulations. High-level data products comprise source catalogs, a variety of sky maps for both gamma-rays and cosmic rays. The highest-level products consist of scholarly papers published in the scientific literature.

Data Format: Raw and reconstructed data from HAWC are stored in a XCDF binary data format developed at UMD. The format is optimized to minimize the data size while not sacrificing the speed of reading the data. The data format and associated tools are freely available under an open source license. Tools have been developed to convert this format to ROOT files. Sky maps will be provided in FITS or a similar format accepted by the community, and other products will be provided adequate format (e.g. ROOT, YAML, HDF5). Plots, presentations, and papers will be stored in PDF format.

Archiving of Data: All HAWC data is archived both at the University of Maryland and at UNAM in Mexico. It is expected that each data archive will contain about 700TB raw of data and about 100TB of data from other sources. including reconstructed data and simulations, for each year of HAWC operations and ideally the archive will be maintained for at least a decade after HAWC is decommissioned.

Metadata: There is no standard metadata format for the field, but the HAWC collaboration has developed databases that track the calibration and data quality of the dataset at each stage of processing. Data archive directory metadata will be used to synchronize data between the two data archives at UMD and UNAM.

Access to Data and Data Sharing Practices and Policies: This project will conform to all HAWC Collaboration policies regarding access to raw HAWC data. Formal policies are described in the collaboration charter. Access to raw data will require that a research collaboration of some sort be formed or worked out on request through consultation with the collaboration. Note that the size of the anticipated 10-year data set for the entire HAWC detector will be approximately 8 Petabytes.

HAWC will provide prompt transient alerts to the Gamma-ray Coordinates Network (GCN), the Astrophysical Multi-messenger Observatory Network (AMON) and IAU Astronomer's Telegrams (ATel). We will provide yearly gamma-ray and cosmic-ray sky maps in FITS format over the Internet, after a period of vetting and analysis by collaboration members. We have also provided an Internet-based tool to examine the sky maps and extract HAWC source fluxes or flux upper limits for all areas of the sky.

Policies for Re-Use, Re-Distribution, and Production of Derivatives: This project will conform to all HAWC Collaboration policies on the use of HAWC data. It is likely that re-use will require an acknowledgement, in published results, of the HAWC Collaboration and its funding agencies.

Data Archiving and Preservation: The HAWC data is stored on disk at 2 mirrored locations (UNAM in Mexico and UMD in the US). Significant subsets are also backed up onto google Cloud drives to maintain long-term data integrity. Upon the completion of the project, the data will be permanently archived by the participating universities.

Public Data Release: The HAWC Collaboration maintains a website with public data releases of high-level data. This website is <https://data.hawc-observatory.org/datasets.php>. As of November 2018, this contains:

1. 2HWC Survey - This provides access to the HAWC data based on 507 days of operation and provides information on the maximum steady TeV flux across more than 8 sr of the sky. \sqrt{TS} maps are available for all locations in the sky visible to HAWC are available, whether or not they are part of the catalog. We provide significance, best flux estimate and a flux confidence interval.
2. Geminga Paper - This page contains the data and code to reproduce the results presented in the 2017 HAWC Science paper Extended TeV Gamma-Ray Sources Around Pulsars Constrain the Origin of the Positron Flux at Earth paper.
3. SS-433 - This page contains the data and code to reproduce the results presented in the 2018 HAWC Nature paper: A.U. Abeysekara et al., Nature (2018) Very high energy particle acceleration powered by the jets of the microquasar SS 433.
4. Lightcurves - This data set contains daily light curves for the Crab Nebula and the BL Lac objects Markarian 421 and Markarian 501 from HAWC observations between 2014 November 26 and 2016 April 20.

We are making plans for a full public release of HAWC data and the software tools required to analyze it in order to provide members of the greater gamma-ray community the ability to fit HAWC data to arbitrary spectral and morphological hypotheses and combine HAWC data products with Fermi and other experiments.